A Quality of Experience Model for Haptic Virtual Environments

ABDELWAHAB HAMAM and ABDULMOTALEB EL SADDIK, University of Ottawa
JIHAD ALJA'AM, University of Qatar

Haptic-based Virtual Reality (VR) applications have many merits. What is still obscure, from the designer's perspective of these applications, is the experience the users will undergo when they use the VR system. Quality of Experience (QoE) is an evaluation metric from the user's perspective that unfortunately has received limited attention from the research community. Assessing the QoE of VR applications reflects the amount of overall satisfaction and benefits gained from the application in addition to laying the foundation for ideal user-centric design in the future. In this article, we propose a taxonomy for the evaluation of QoE for multimedia applications and in particular VR applications. We model this taxonomy using a Fuzzy logic Inference System (FIS) to quantitatively measure the QoE of haptic virtual environments. We build and test our FIS by conducting a users' study analysis to evaluate the QoE of a haptic game application. Our results demonstrate that the proposed FIS model reflects the user's estimation of the application's quality significantly with low error and hence is suited for QoE evaluation.

Categories and Subject Descriptors: H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems—*Evaluation/methodology*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*User-centered design*; H.1.2 [**Models and Principles**]: User Machine Systems—*Human factors*

General Terms: Human Factors

Additional Key Words and Phrases: Haptic I/O, quality of experience, QoE, virtual reality, haptic multimedia application, fuzzy logic evaluation

ACM Reference Format:

Abdelwahab Hamam, Jihad Alja'am, and Abdulmotaleb El Saddik. 2014. A quality of experience model for haptic virtual environments. ACM Trans. Multimedia Comput. Commun. Appl. 10, 3, Article 28 (April 2014), 23 pages.
DOI: http://dx.doi.org/10.1145/2540991

1. INTRODUCTION

Traditionally multimedia systems involve the use of multiple forms of information contents such as graphics, images, 3D models, and audio/video files. Recently, multimedia research has expanded beyond the traditional focus on audio and video media to encompass other media of growing importance; in particular haptics or the sense of touch. The incorporation of the sense of touch gives rise to far

This publication was made possible by a grant from the Qatar National Research Fund under its award NPRP 09-052-5-003. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Qatar National Research Fund.

Authors' addresses: A. Hamam and A. El Saddik, Multimedia Communications Research Laboratory, School of Electrical Engineering and Computer Science, University of Ottawa, 800 Avenue King Edward, Ottawa, Ontario, K1N 6N5, Canada; email: elsaddik@uottawa.ca; J. Alja'am, Department of Computer Science and Engineering, University of Qatar, Qatar.

DOI: http://dx.doi.org/10.1145/2540991

more exciting and appealing ways of supporting collaboration, copresence, and togetherness in these multimedia systems by enabling users to feel each other's presence and the environments in which they are interacting [El Saddik 2007].

Although the advantages of haptic virtual reality (VR) applications are promising, it is still not vividly determined how the users will experience those applications. There is a lack of objective measurement of these advantages through a concrete evaluation model. Quality of Experience (QoE) for haptic virtual environments is an evolving research topic concerned with evaluating the user experience with the multi-modality application. QoE is an indicator of the quality of involvement and interaction between the user and the designed application, rather than an indication of functionality and the quality of the technology used in the application.

QoE is more than just assessing the Quality of Service (QoS) that an application provides to users [Jain 2004]. While QoS is part of the assessment, whether it is jitter and delay of the network or intermodal/intramodal synchronization of multimedia contents, there are still other parameters to consider such as ease of use, rendering quality, and measurement of fatigue. These parameters represent the user experience and are subjective in nature. Both the QoS and the user experience compose the overall QoE which in turn will reflect the perceptual value of multimedia applications [Jain 2004].

The future of multimedia design metrics will eventually be based on QoE assessment rather than QoS assessment. In their 2005 multimedia retreat report, Rowe and Jain [2005] indicate that users will be the focus of assessment and their experience is the basis of building future multimedia applications. There are challenges arising, though, that revolve around the subjectivity of the user, the diversity of the applications, and the variation of the environmental context [Jain 2004]. Eventually, users are going to be working with the application and their perceived experience is the ultimate goal.

This article proposes a taxonomy for QoE evaluation metrics associated with Haptic-Audio-Visual Environments (HAVE). This taxonomy includes an extensive list of related parameters which are contributing factors for the assessment and testing of the quality perspective of users with HAVE applications. Furthermore, parameters from the proposed taxonomy are modeled using the Mamdani Fuzzy logic Inference System (FIS). Users' studies were conducted at the DISCOVER lab, University of Ottawa to assess a haptic-enabled VR application using the FIS QoE model and compare those results with the overall users' self-assessment of the application using statistical analysis techniques.

The rest of the article is organized as follows. Section 2 reviews the related work in the field of QoE for multimedia and VR applications. In Section 3, we present the taxonomy for QoE parameters including the complete charts and the rationale behind that taxonomy. Next, the fuzzy logic model is proposed in Section 4 along with the result of the users' study. In Section 5, we test the FIS model and analyze the results using error and statistical testing. Finally, we conclude this article in Section 6.

2. RELATED WORK

It has been widely accepted that multimedia research should focus on QoE as the primary metric for evaluating the user experience in multimedia applications; in contrast to QoS [Jain 2004; De Marez and De Moor 2007; Wu et al. 2009; Ebrahimi 2009]. Research that incorporates the user is more challenging because human behavior is so variable. Nevertheless, the goal of nearly all multimedia applications is to solve a problem for a user, so user perception must be at the core in an evaluation metric for any desired algorithm or application.

In VR, some authors [Rowe and Jain 2005; Ebrahimi 2009] stress the move from QoS metric to QoE for newer multimedia data types and human computer interaction, which could be the future of interaction. In their research paper "Quality of Experience in Virtual Environment," Gaggioli et al. [2003] look at experience from a holistic view. They first argue that some authors equate virtual experience with presence which is biased since many other elements are included in the human experience such as emotions and enjoyment. They define what is called optimal experience that is a state of consciousness having positive, complex, and rewarding properties. This optimal experience is a combination of genetics and innate human qualities. VR has the four characteristics necessary for creating optimal experience: opportunities for action, skills, feedback, and control.

2.1 Measuring QoE in Virtual Reality

Measuring the QoE requires measuring the user's perception of the VR application. As Jain puts it [Jain 2004], we require improved performance measures over the well-established QoS measures to deal with the subjectivity of the user. Whalen et al. [2003] discuss some of the methods and challenges in determining those performance measures in the context of VR.

Traditionally, there are three recognized methods for assessing the user's feedback and responses in a virtual environment (VE): subjective, performance-based, and physiological measures [Whalen et al. 2003]. Each method enables the collection of a specific type of information regarding the user's responses to the application. For instance, subjective measures evaluate the user's satisfaction, fatigue, intuitiveness, preferences, etc. (collected via surveys). Performance measures evaluate the user's behavior when performing a task with the VR application. Finally, the physiological measures evaluate nonvoluntary responses of the human body during and immediately after the test session.

One research group [Huang et al. 2012; Iwata et al. 2010; Tatematsu et al. 2010; Kusunose et al. 2010; Ida et al. 2010] relies on subjective measures solely to assess the QoE of various applications. They use Mean Opinion Score (MOS) to determine the overall QoE. Their main idea is to test the network jitter and delay effect on the haptic quality reflected on the user. Using different settings and jitter parameters they test different scenarios using a networked haptic application. In one scenario they test a hockey game, while in another they test a networked writing application. One of their papers assesses haptic and video and audio, another focuses on haptic, while yet another focuses on the quality of a first person shooter game. From the MOS values, the research group uses multiple regression analysis to link QoE parameters with application-level parameters.

Video playback using haptic feedback evaluation was performed in Danieau et al. [2012], where the user is presented with a questionnaire after using the application. The QoE is computed as a sum of four constructs (realism, sensory, comfort, and satisfaction). Three modes of the application were compared (feedback, random feedback, and no feedback), in which realistic feedback gained the highest QoE. Tactile and kinesthetic haptic devices were found to increase the QoE of the application in Hamam et al. [2013]. The method used was a subjective questionnaire in which the user evaluates an application with and without the presence of haptic feedback.

In Hamam and El Saddik [2013], subjective assessments of VR parameters are used to compute the QoE of haptic-based VR applications. The computation is done through a mathematical model consisting of three tiers of formulas to compute the QoE through weighted averages. The paper discusses ways to determine and calibrate the weights of the formulas through five techniques: even weights, correlation based weights, even-weights/correlation amalgamation, linear regression, and principle component analysis. The QoE value generated with each technique is analyzed and compared using statistical analysis.

Subjective and performance-based QoE evaluation research has been performed in the haptic field. Nonetheless, the evaluation methods and the aspects to be evaluated vary depending on the type of the application and the parameters to be evaluated. For instance, Basdogan et al. [2000] combine both types in their studies to evaluate the incorporation of the haptic media in collaborative human-human and human-machine interactions in shared virtual environments (SVEs). The evaluation consisted of a measurement of response variables as well as conducting a questionnaire to the users undergoing the experiment. A similar approach that measures haptic benefits in SVEs is presented in Guerraz et al. [2003]. The authors measure physical parameters generated by the haptic device directly in order to assess the quality of the application. The authors suggest that this is an assistive approach to conducting a statistical survey. Examples of parameters that are measured and included in the physical portion are gesture position and gesture velocity, where a gesture is defined as the way the user manipulates the haptic device.

Performance-based QoE can be often interchanged with Quality of Performance (QoPer) which is another user-centric quality metric [Roid 2004]. QoPer methods refer to observational procedures or testing, that are designed to evaluate the correctness of performing a particular task or how well the user performs. Compared to the QoPer approach that is task and procedure oriented, the QoE approach measures the implications of the interaction at the cognitive level (which is why it is often accompanied by a questionnaire).

QoPer measures have been used in VE research. For instance, a quality of performance model is proposed in Alamri et al. [2008] to evaluate a stroke patient's performance in a rehabilitation system. The authors propose an evaluation taxonomy that includes the following factors: task completion time, eye-hand coordination, compactness of task, hand movement and steadiness, grasping angles, and fingers grip accelerations. Another work [Park and Kenyon 1999] studies the effect of network latency and jitter on performance in a collaborative VE, where two subjects manipulate a ring along a complex trajectory. The performance measures include the subjects' speed (time to complete the task) and error rate (number of collisions between the ring and the path).

Some research efforts utilize physiological evaluation of VR systems. [Jones and Ho 2008] discuss the evaluation of thermal displays using two approaches: physiological approach through thermal sensors and a psychophysical approach. Karam et al. [2009] state their intention to add biometric measures for physiological evaluation of an audio-tactile display in the future (such as respiration and heart rate) to reinforce the user's evaluation. Specific to the QoE, Whalen et al. [2003] suggest the use of physiological measures to determine the QoE of VR applications. With stress as an example, the authors argue that the sympathetic nervous system is activated and blood volume, heart rate, and respiration rate all increase. By measuring these parameters, an estimation of the stress level can be made.

When dealing with cybersickeness resulting from VR environments, Ramsey [1997] claims that measuring physiological symptoms directly is more effective than a questionnaire due to three limitations:

—People are mentally aware of their internal state (emotional condition) when under the same circumstances, in the real world, they would not normally be.
—People might not understand the implication of the response in the questionnaire.
—People may not wish to report feeling any symptoms of sickness.

There is a continuing research effort in psychophysiology that tends to evaluate elements of the user experience by subjective and objective measures. The research focuses on the users' emotions during the interaction with an environment such as the playing of a video game. The subjective measures would be the user emotions assessed via questionnaires while the objective measures are the corresponding physiological measurements of the user.

For example Nacke and Lindley [2008] establish the correlation value between the subjective and objective indicator of the experience during the gameplay of a first-person shooter game. The correlation value examines the link between the physiological measure and the corresponding psychological measure.

2.2 QoE Frameworks

[Wu et al. 2009] have developed a conceptual framework of QoE for disturbed interactive multimedia environments. The framework divides the user experience into cognitive perception and behavioral consequences. They model the QoE based on parameters from both categories of the user experience. Cognitive perception impacts the user's behavior, whether it is subjective or objective responses of the user. Most importantly, QoS is modeled as a separate construct than the QoE construct. Both constructs are linked through a causal chain where QoS drives the QoE of the user. QoS parameters, quantifying the environmental influences (where the environment is the technological system), are linked to the cognitive perceptions and behavioral consequences of the user. Through controlled and uncontrolled studies, the authors map QoS with QoE by quantifying the correlation between both parameters. The focus in their paper was on technological influences of the environment.

The same framework was used to study the user experience in a tele-immersive gaming system [Wu et al. 2010]. In this study some nontechnical factors were included with the technical ones introduced in the previous study [Wu et al. 2009] in order to address three points. The first point is the people reaction to the new 3D gaming experience. In addition, the impact of QoS factors on QoE in 3D tele-immersive gaming is studied. Finally, other factors effect (age, social interaction, and physical setup) on QoE is explored by getting the users' verbal feedback.

A novel way of evaluating QoE is presented in Bhattacharya et al. [2011]. The authors address the shortcomings that may arise from the adopted models and techniques in evaluating QoE. They propose evaluating QoE through affective computing. Affective computing deals with the analysis of human emotional variables during the interaction between the user and the system. Those emotional variables have a strong association with QoE and they include facial expression, speech, and body gesture. A study is conducted to investigate the hypothesis that affective behavior is correlated with QoE.

2.3 Main Contributions of our Work

We present a fuzzy logic system for evaluating the QoE of VR applications tailored towards the user. This article focuses on the overall QoE that the user undergoes when interacting with a VR application. The main contributions are the taxonomy modified from the previous publication [Hamam et al. 2008a], and a way of assessing QoE using fuzzy logic theory given that the parameters are mostly subjective in nature. Our work lays the foundation for establishing an alternative method capable of evaluating the QoE of haptic-based VR applications without ultimately resorting to users' subjective tests which are time and resource consuming.

Our previous work [Hamam et al. 2008b] relied on expert rules for developing the elements of the fuzzy logic system. However, as we adapt a completely user centric evaluation methodology, we build the fuzzy logic elements in this article from the users data. We divide the data into two sets, one for building the system and the other for testing the system. Then we perform the analysis necessary to verify that the system significantly represents the user. Hence, this current system completely represents the user, and can be regarded as a first step into user replacement tactic for application evaluation that might take place in the future.

3. QUALITY OF EXPERIENCE TAXONOMY

In order to classify different QoE parameters, we first need to define the QoE. Some researchers base their QoE construct on QoS parameters solely. Calculating, estimating, or predicting the QoE from the QoS is very convenient for the developers or evaluators of the application but it deviates from the uniqueness of the QoE construct, which should not be based solely on QoS, in our opinion. This might defeat the purpose of having a separate construct. We acknowledge that QoS should be a part of QoE, as it affects the experience of the user. However, it should be one aspect of that construct among other aspects.

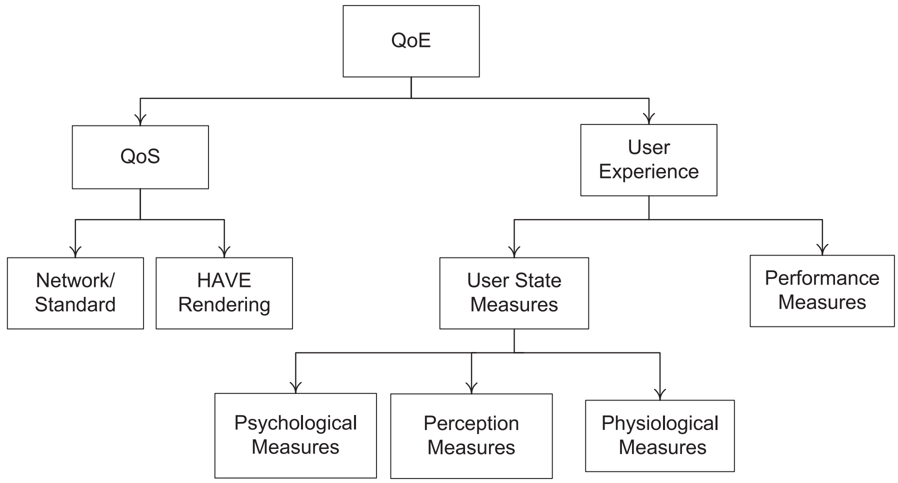

Fig. 1. Higher-level organization of QoE taxonomy.

The definition we have adopted is that the QoE is composed of the QoS and User Experience (UX). This definition would satisfy the properties of QoE being user centric, subjective, and revolving around the experience of the user. Concurrently, the definition of QoE includes the QoS metric. Hence, QoE would not be based only on QoS metric but rather an amalgamation of QoS and UX.

The taxonomy of parameters should stem from the core definition of QoE. Hence, we based our top organization into two parts: QoS and UX parameters. We further subdivided the QoS category into two parts: network/standard parameters and HAVE multimodality rendering parameters. The UX category was subdivided into two parts as well: user state measures and performance measures. In turn, user state measures were divided into different categories as will be detailed later. This higher level organization, shown in Figure 1, reflects an apparent taxonomy for VR applications evaluation and at the same time is more customizable depending on the parameters needed for evaluation. As an example, developers wishing to evaluate only the QoS of the application can disregard the user experience parameters and just focus on QoS.

3.1 Quality of Service Parameters

QoS parameters insure the smooth flow of the application for the user or customer. We have divided the QoS parameters into two parts: Network-related and HAVE modality rendering quality.

3.1.1 *Network/Standard QoS.* Most parameters shown in Table I are standard for any networked application but the parameters might be altered for VR. For example, if we consider synchronization it can be divided into two parts: network synchronization which is common to network applications and media synchronization which is specific to the multimodal side of VE in essence the synchronization of the three media streams: graphics, audio, and haptics.

3.1.2 *HAVE Rendering Quality.* The rendering quality relates to the quality of the three major modalities in VR applications, namely: graphics, audio, and haptics. Each modality is considered separately first and eventually blended and mixed modalities are considered. As seen in Figure 2, there is an emphasis on the haptics modality as it is the newest modality to be introduced in VR applications, and it has very stringent requirements in terms of feedback loops which might affect the stability and transparency of the application.

Table I. Network/Standard QoS Parameters

Response Time	The time taken by a system to respond to an action, measured in millisecond or microsecond
Latency/Delay	Time taken for the packet to reach from source to destination, measured in millisecond or microsecond. There are different source of delay. These are 1) Propagation Delay which is the delay through a physical medium, (2) Queuing delay which represent the time spent in router queues and (3) Hop Count where each traversed router or switch adds queuing delay to the overall route
Price	The quantity of payment or compensation given from one party to another in return for goods or services, it can be measured by a metric related to energy, monetary, automation or other efficiency of the service.
Privacy	Deals with what personal information can be shared with whom and whether messages can be exchanged without anyone else seeing them
Security	The level of protecting the information exchanged through the use of multimedia technology
Availability	The ratio (or probability) of time a system or component is functional to the total time it is required or expected to function. Small probability values for availability indicate bad QoS, while high values indicate good QoS.
Bandwidth/Throughput	The amount of data transferred from source to destination or processed in a given amount of time. Measured typically in bits/second or bytes/second
Synchronization: Network Synchronization (CVE) Media Synchronization (intramodal)	Network Synchronization: the temporal relations linking the various media objects within a multimedia presentation. Example: Time relations of a multimedia synchronization that starts with an audio/video sequence, followed by several pictures and an animation that is commented by an audio sequence and haptics feeling. Media Synchronization (intramodal): Refers to the temporal relations between media units within a time-dependent media object. For a video with a rate of 25 frames per second each of the frames has to be displayed for 40 msec. For haptic data with 1 KHz, each of the data samples must be captured and displayed for 1 msec
Jitter	Difference in latency of network packets usually measured in microseconds or nanoseconds
Reliability	The ability of the computer system and its components, i.e., the haptic audio visual environment to consistently perform according to the given specifications
Error Magnitude Frequency	Sometimes CHAVE packets are corrupted due to bit errors caused by noise and interference. The receiver has to detect this and, in case the data contained in the packet is needed, may ask for this information to be retransmitted.
Safety	The needed aspects to be considered in order to operate the haptics environment properly and use it in conjunction with other peripheral equipment without damaging the environment and the users

3.2 User Experience

The second part that constitutes the definition of QoE is the UX. This is an important evaluation category for the overall quality of the application. Even if the application possessed excellent QoS parameters still users might feel that the application is not up to their expectations for some reason. For instance, the application might not be exciting enough, difficult to use, or causing dizziness which is referred to as cybersickness.

3.2.1 *User State Measures.* The first subcategory of the UX reflects the emerging state of the user. The perception, psychological behavior, and physiological symptoms of the user are all present in this

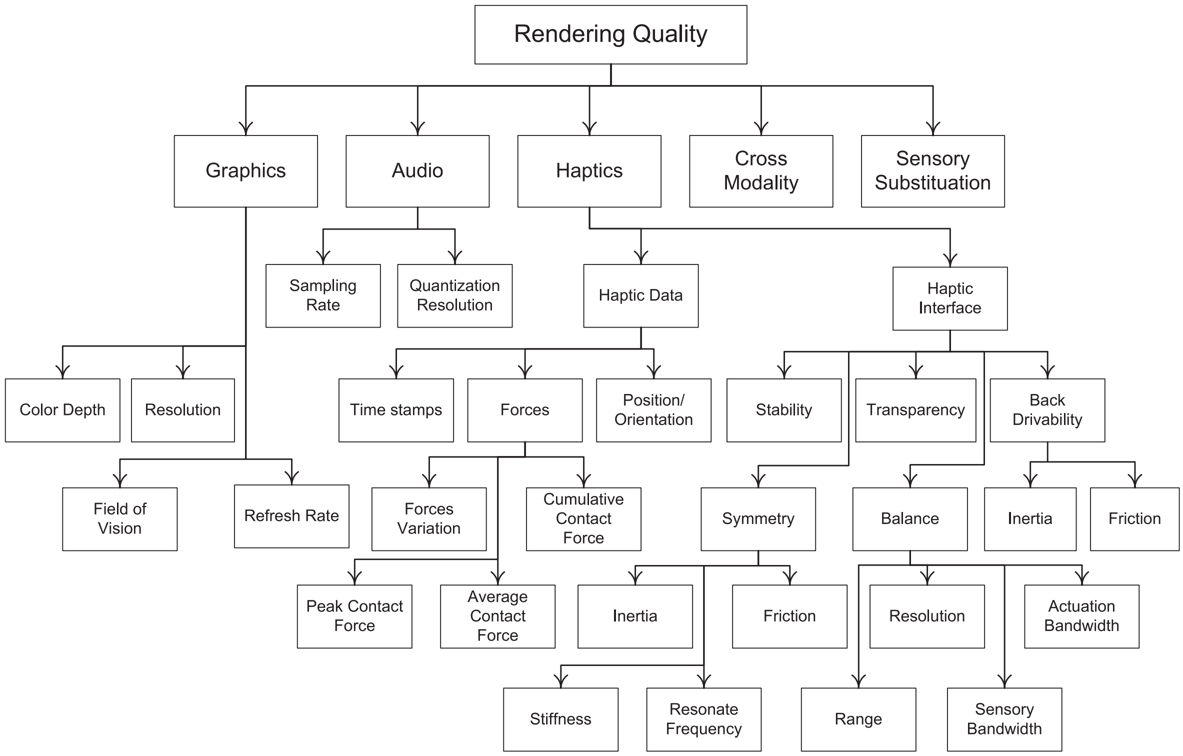

Fig. 2. Parameters of rendering quality.

category. The user state group is subdivided into three measures: perception, psychological, and physiological. Two of these groups (perception and psychological) embodies the subjective elements of the user state while the physiological measures stand for the biological elements of the user state.

3.2.1.1 *Perception measures.* As depicted in Figure 3, perception measures mirror how the user perceives the application. This is a user-centric category, and could be unique for every user. Some users may get tired from the application, while others may feel relaxed. Some might feel the effect of collaboration in a collaborative Haptic Audio Visual Environment (c-HAVE) while others might need more stimuli. Each user may have a certain set of preferences and modality choice given the versatility of human nature. The perception group reflects on the user perceived state mainly through questionnaires. We have tailored the perception measures around the three pillars of VR perception: presence, immersion/involvement, and engagement/enjoyment/flow.

3.2.1.2 *Psychological measures.* The next category which constitutes the other subjective part of the user state category is psychological measures. Psychological measures reflect the psychological state of the user through observation and user feedback but not direct physiological measurements. Observation can assess the psychological behavior of users, such as stress, without hindering the user movements by including measuring devices. Psychological measures are displayed in Figure 4.

In Figure 4, emotions are subdivided into two parts: negative emotions and positive emotions. These terms are relative to the user and not the experience. Negative emotions can produce undesired responses if the premise of the application did not intend to produce these emotions. Even users' positive emotions are capable of producing undesirable effects but to a lesser degree. For example, there

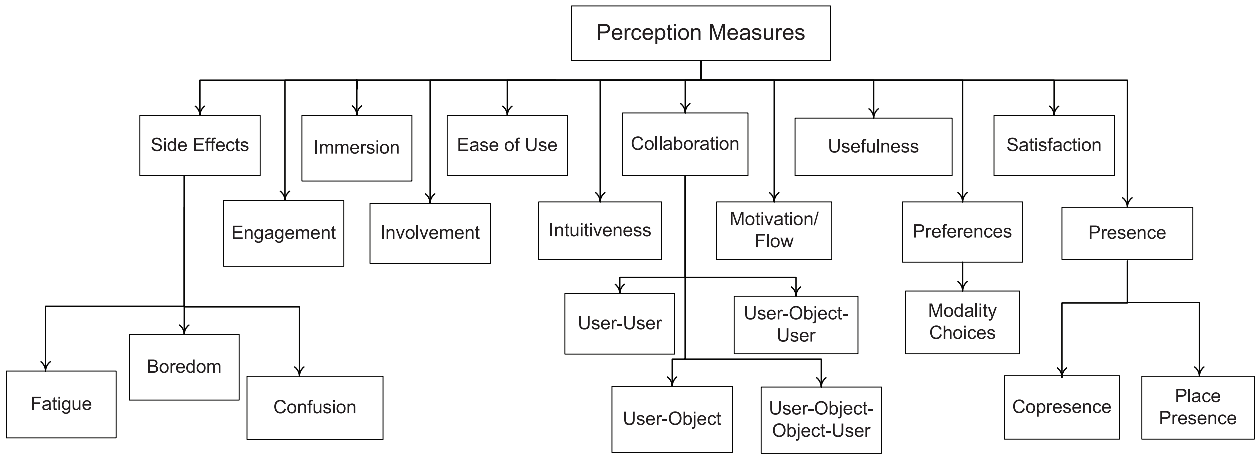

Fig. 3. Parameters of perception measures.

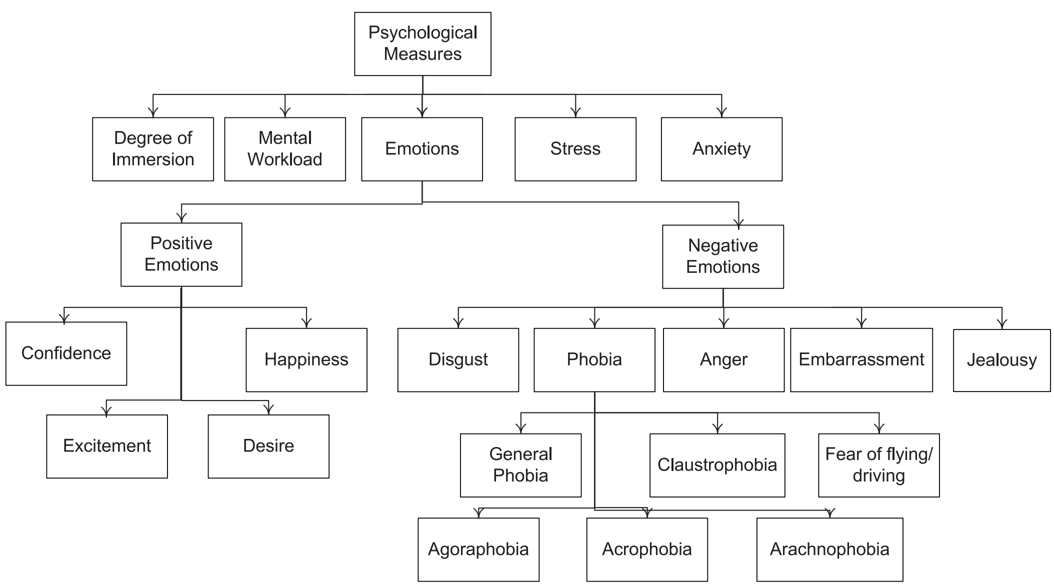

Fig. 4. Parameters of psychological measures.

is the difficulty of reentering into the real world after a joyful and pleasant experience in the virtual world. This turbulence of emotions may decrease the user's experience if not handled gracefully [Behr et al. 2005].

Although immersion is included in the perception measures, the degree of immersion parameter is included in this category. This is because immersion has two definitions in the literature. The first definition regards immersion as a perception measure and the other definition as a property of the environment. For the latter case degree of immersion would be a psychological response of the user. For definition of the phobia terms please refer to Strickland et al. [1997]. Strickland et al. discuss VR

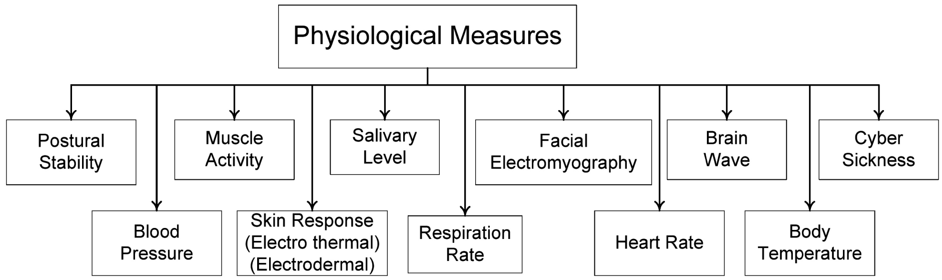

Fig. 5. Parameters of physiological measures.

applications that can cure individuals suffering from phobia. Nonetheless, if the application's goal is not to address phobia, any fear influence will decrease the QoE considerably.

3.2.1.3 *Physiological measures.* The third set of the user state measures indicates the state of the user through biological means. Unlike perception and psychological measures which are assessed through questionnaires and observations, physiological measures, shown in Figure 5, are biological parameters measured directly through users while they are using the application. These parameters indicate relevant factors such as cybersickness, stress, and brain activity [Whalen et al. 2003].

As mentioned in the related work, some research efforts are linking subjective user state parameters with objective ones. For example, Nacke and Lindley [2010] focus on immersion, flow, boredom, and presence as subjective user state parameters. They link those parameters with physiological parameters such as electromyography and electrodermal activity during the user gameplay.

3.2.2 *Performance Measures.* Performance measures govern how the user performs using the VR environment. Currently there are three parameters that falls under this category: task completion time (TCT), task learning time, and error rate. This category is application dependent and developers may wish to alter the difficulty of the application (such as games). This will increase the performance measures parameters but not necessary decrease the QoE.

4. FUZZY INFERENCE SYSTEM EVALUATION

Fuzzy inference system (FIS) maps input values to an output value using fuzzy logic. The input-output mapping is done by evaluating a set of rules (if-then statements) provided to the system by experts or user testing results. Each input or output parameter is defined by a fuzzy set. Fuzzy sets are sets without a clearly defined boundary, for example the set of tall people. Fuzzy sets can be described by curves referred to as membership functions (MFs). A membership function is curve that maps a point in the input space to a membership degree that specify how strongly this input point belong to that particular fuzzy set. Going back to the example of tall people, a 7 feet input would have a very strong degree of membership within the set of tall people [MATLAB Documentation 2001].

As can be observed from Section 3, many QoE parameters are subjective and are fuzzy in nature. For instance there is no crisp answer to whether the user is under stress or whether the application is easy to use. Therefore, a fuzzy logic system is needed to map the fuzzy logic inputs to a crisp output, which is in our case a QoE value. The system would vary in the number of inputs provided along with their MFs, depending on the type of application we are trying to evaluate [Hamam et al. 2008b]. As a proof of concept, we picked out five parameters that are relevant to a particular collaborative VE application, named Balance Ball game, where the user is immersed in a 3D application. The five parameters act as the input to the FIS as described in the following subsections: selecting the relevant parameters for the

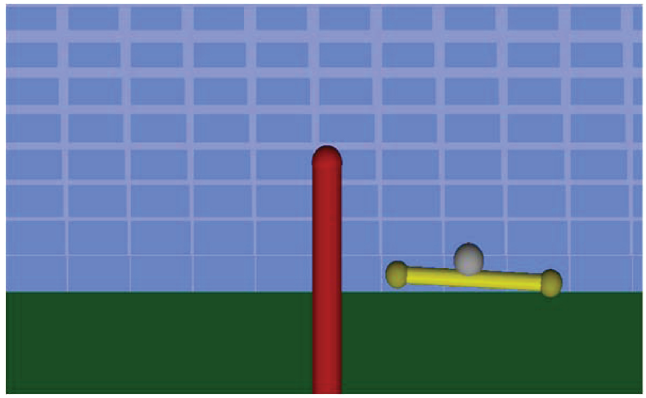

Fig. 6. The “Balance Ball” game snapshot.

haptic application, selecting the FIS to be used in the evaluation, designing the fuzzy sets, generating the fuzzy rules, and finally defuzzification of the output to get a crisp QoE value. But first, we begin by describing the application and the user experimentation conducted to generate the results.

4.1 Application Description/Experimental Setup

The application we used to test the proposed model is the Balance Ball game [Al-Osman et al. 2008], shown in Figure 6. A ball is placed on a long wooden board that is held by two players, one from each side. The game involves the two users to collaborate in order to maintain the balance of a virtual ball on a board using remote haptic devices. Each player holds one end of the board with his/her haptic device and raises it slowly over a virtual pole to a predefined end mark. The challenge is to collaborate in an attempt to keep the board horizontally balanced as much as possible from the initial location to the destination. Any variation in the horizontal balance will cause the ball to roll away towards one side thus penalizing both players. The players should remedy that by using the force feedback and the 3D graphics to apply their judgment in balancing the board again. The score consists of the task completion time and the variations of the ball’s position from the middle of the board.

The experiment took place at the haptic laboratory of the DISCOVER Lab at the University of Ottawa. Thirty users participated in the experiment. The collaborative application ran on two computers. Each user was situated on a separate computer without the ability to view his/her coplayer screen. The computers were running WinXP SP3 on a 2x Intel Xeon 2.8GHz with 2GB of RAM and an Nvidia QuadroFX 2000XGL 128Mb DDR video card. A Phantom Desktop haptic device was attached to it. The Phantom Desktop is a six degrees of freedom positioning and sensing haptic device developed and marketed by SensAble Technologies Inc. It has a compact design and provides three degrees of freedom output capabilities. A snapshot of the experiment setup is shown in Figure 7.

Users were selected randomly from the School of Electrical Engineering and Computer Science (EECS) department at the University of Ottawa (12 females and 18 males, the age is estimated to be between 20–35 years old as age was optionally provided by the users). Twenty users were familiar with haptic devices (among them eleven users with previous working haptic experience), while the other ten users were new to the haptic notion. In either case, the user was given a general background about the application, how to handle and hold the haptic device, and the goals of the experiment. Users were reminded that the purpose of the experiment is to evaluate the application and not the users’ abilities.

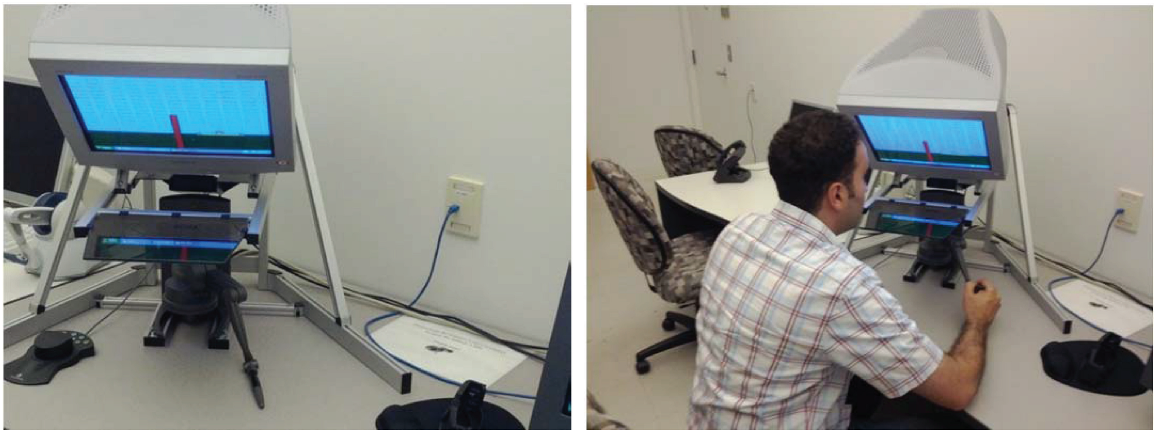

Fig. 7. The experimental setup of the user study conducted. The figure displays the station setup of one of the users during their gameplay.

Users were divided into teams. Each team consisted of two users, and the experiment lasted around fifteen minutes on average, which included playing the game twice (the first time was a trial game, while the second one was the actual game). After the team finished playing the virtual game collaboratively by reaching their destination in the actual game, they were asked to fill out a questionnaire with general questions about the virtual game, past haptic experience, and specific questions that reflect elements of their experience. Each user filled out a separate questionnaire.

4.1.1 *Results.* The format of the questionnaire and the questions included can be viewed in the Appendix. We have used some of the users' results to build the FIS while other results were used for testing. The results were converted to a normalized percentage form (this was a necessary step to build a standard membership function as described later). The questionnaire presented to the users was a five point likert scale in which each question is followed by two anchor labels with five points to choose in between. Hence, each user selected a value from one to five. To convert this value into a percentage, a formula is given by Preston and Colman [2006]. The formula is

$$(\text{rating} - 1)/(\text{number of response categories} - 1) \times 100. \quad (1)$$

Hence after the conversion, the raw data became: 1 −> 0, 2 −> 25, 3 −> 50, 4 −> 75 and 5 −> 100. The overall rating is the user's perceived QoE and it is given by the user as a percentage form, no conversion was necessary in that case.

The users for testing were picked randomly by choosing every 6th user out of 30 users (the 6(*n)th users, $1 \leq n \leq 5$). The rest of the users' results were used to build the FIS. Systematic sampling, which is employed here, is used in statistical psychology and is valid as long as the data are randomly organized [Black 2004]. In our case there was a random organization of the data where each user had a random chance of being the nth user in the list.

4.2 Selecting the Input Variables

Five parameters were selected to act as input to our FIS. They are listed here along with a description of each parameter as well as the reason why it was selected.

(1) *Media synchronization (Network QoS parameter)*. There are usually three media modalities in a HAVE application. Any miss-synchronization between the audio, graphics, and haptics can cause

a loss of perception of the media included in the application. Therefore media synchronization is necessary for players to maximize their perception and enjoy the game. In this particular case, we focus on the subjective aspect of media synchronization from the user's point of view (even though it can be analyzed through equations our focus is the user's perception and experience).

(2) *Fatigue (perception measures parameter)*. Research has shown that fatigue, which is caused by muscle exhaustion, is linearly distributed as a function of time [Seroussi et al. 1989]. Fatigue is a crucial parameter because the haptic application needs users to interact with the virtual environment by exerting force and this induces fatigue easily compared to audio-visual feedback. Depending on the specifics of an application and on the haptic device used, rapid fatigue can hinder users and limit their rapport with the application. On the other hand if the application minimized users' fatigue then their experience will be more positive.

(3) *Haptic rendering (QoS HAVE rendering quality parameter)*. Haptic rendering quality remains the same until we reach a threshold (that is usually referred to as the JND: Just Noticeable Difference) after which the quality starts decaying [Srinivasan and Basdogan 1997]. For any haptic application, we want the quality to remain above that threshold, otherwise any instability, low resolution, or low haptic fidelity will render the application virtually unrealistic from the user's point of view.

(4) *Degree of immersion (psychological measures parameter)*. Even though the degree of immersion will cause a difference in quality, this difference is still not quite understood [Gutierrez et al. 2007]. However immersion in gaming applications is of importance, since the more the users are immersed in the game the more they are involved and experiencing enjoyment [Ijsselsteijn et al. 2007].

(5) *User intuitiveness (perception measures parameter)*. User intuitiveness is an important phenomenon that has been considered in disciplines other than human-computer interaction, such as nursing [Miller 1993]. Although the factors that contribute to intuitiveness are little known, it can be observed through swift and determined actions of the user. It can be determined through user feedback as well.

The selection of the parameters is tailored towards the new medium experience and game experience holistic theory. The new medium (haptic) is emphasized by the two parameters: media (graphic and haptic) synchronization and the haptic rendering parameter that is the top level parameter of the haptic node under the HAVE rendering quality measures in Figure 2.

Game experience psychologists have divided the experience for users of digital games into two categories; immersion and flow [Ijsselsteijn et al. 2007]. Under flow, there are certain characteristics that model an acceptable level of enjoyment for the user. Most importantly, the interface should not be too cumbersome and it should be responsive to the user (intuitiveness). In addition, the lack of fatigue will increase the flow and enjoyment level among users [Sweetser and Wyeth 2005].

The parameters selected represent all the categories in the taxonomy except for the performance measures. Our UX focus was on the user state (through perception and psychological measures) rather than performance of the user, because the performance of the user was unrelated in this case, as we had no expectation for performance. We wanted to focus on the user state parameters and the effect of the HAVE application on the user. Since the parameters selected stem from game theory, they are relevant to any haptically rendered three-dimensional collaborative game application such as the Balance Ball game that we tested.

4.3 Selecting the Fuzzy Inference System

Using MATLAB [MATLAB Documentation 2001], we have built a fuzzy logic system to test our QoE model and apply it in evaluating a multi-modal environment. We have implemented the model using

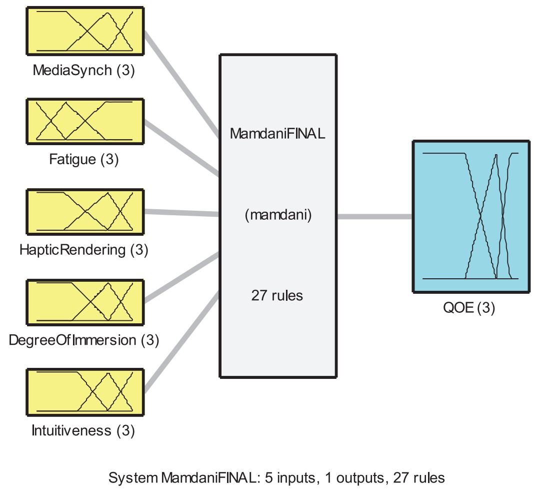

Fig. 8. Mamdani fuzzy inference system.

the well-known and established Mamdani inference system [Mamdani and Assilian 1999]. As seen in Figure 8, the five parameters described above act as inputs to the system. The Mamdani system applies defuzzification to the output, which is as well modeled with membership functions (MFs), to generate a crisp output value. Naturally in our case, this output value is the QoE of the user, based on the inputted parameters.

4.4 Defining Fuzzy Sets (Clustering Technique)

In order to define our fuzzy sets and generate membership functions for our input parameters, we have followed the method described in Hajshirmohammadi and Payandeh [2007]. Using the fuzzy c-means (FCM) clustering technique we were able to generate clusters of data for each parameter and then, based on the result of the clustering, we generated a standardized fuzzy set. We used the data of 25 users to build the FIS system, while five users were used for testing.

FCM is a clustering tool provided to find natural grouping in data to represent a system's behavior. Unlike other clustering techniques, FCM exploits the membership grade to group the data points into clusters. Each data point belongs to a cluster to some degree that is specified by a membership grade, thus enabling the clustering of multidimensional data.

Based on the results of the questionnaire, we created a two-dimensional data vector for each parameter consisting of the index representing the users' cardinal order and the value corresponding to the users' rating of that particular parameter. We used the parameter's data vector as a data-input to generate three clusters for each parameter using FCM. As an example of the clustering results produced, Figure 9 shows the three clusters' centers labeled A, B, and C belonging to the media synchronization parameter.

Many papers that standardized fuzzy sets from training data used trapezoidal and triangular MFs for the input parameters given their simplicity, low computation requirements, and the sufficiency to capture fuzzy modularity [Hajshirmohammadi and Payandeh 2007; De Vleeschouwer et al. 1997; Wang and Mendel 1992]. These common membership functions' shapes were also used in the design of the parameters in our system. For every parameter we followed these guidelines.

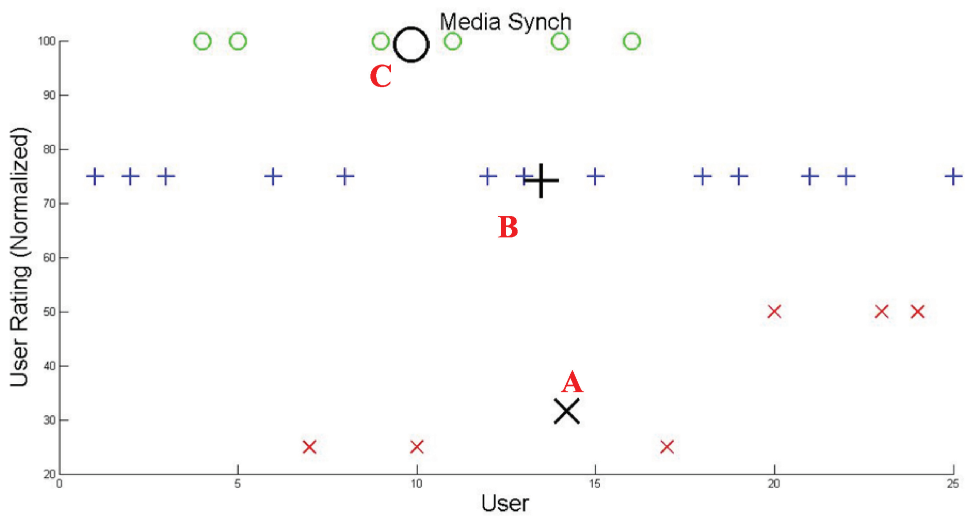

Fig. 9. Clustering results for the media synchronization parameter.

—The first MF is a trapezoidal membership function. It starts with the maximum truth value of one until the value of the first cluster center is encountered. The MF starts decreasing until it reaches a zero truth value at the second cluster center value.

—The second MF which is triangular in shape starts at the value of the first cluster center and ends at the value of the third cluster center. It reaches its maximum truth value of one at the value of the second cluster center.

—The third MF is also trapezoidal. From the value of the second cluster center the function starts rising gradually to the maximum truth value of one at the third cluster center. The MF stays at one till the end of the range of values.

—At each x-axis value, the aggregate truth value of all MFs is one. To that effect the intersection of the MFs occurs at 0.5 truth value. At this point each one of the intersecting MFs has equal truth value of 0.5 and the aggregate is one.

The output parameter, as well, was divided into three MFs. Unlike other FIS systems we looked at in the literature, we wanted to standardize the output MFs too instead of just dividing them into three equally spaced MFs. As a result, we applied our clustering technique and the same set of the above guidelines to the output dataset, which is the overall rating of the users. All the input parameters' MFs and the output parameter's MFs can be viewed in Figure 10.

It can be noted that when building our fuzzy sets we used linguistic labels to name each MF of every parameter. By harnessing the linguistic advantage of fuzzy logic we give meaning to our MFs instead of just labeling them low medium and high. However, in the next section we keep referring to MFs as low, medium, and high to be consistent when discussing the derivation of rules.

4.5 Deriving Fuzzy Rules

Fuzzy rules were derived from the results of the users. We generated the rules based on the results of the twenty five users that were used to build the data clusters. Table II illustrates this with an example of rule generation from the data of user 3. When referring to the rules, the notion L is for the lower MF, M for the middle MF, and H for the high MF.

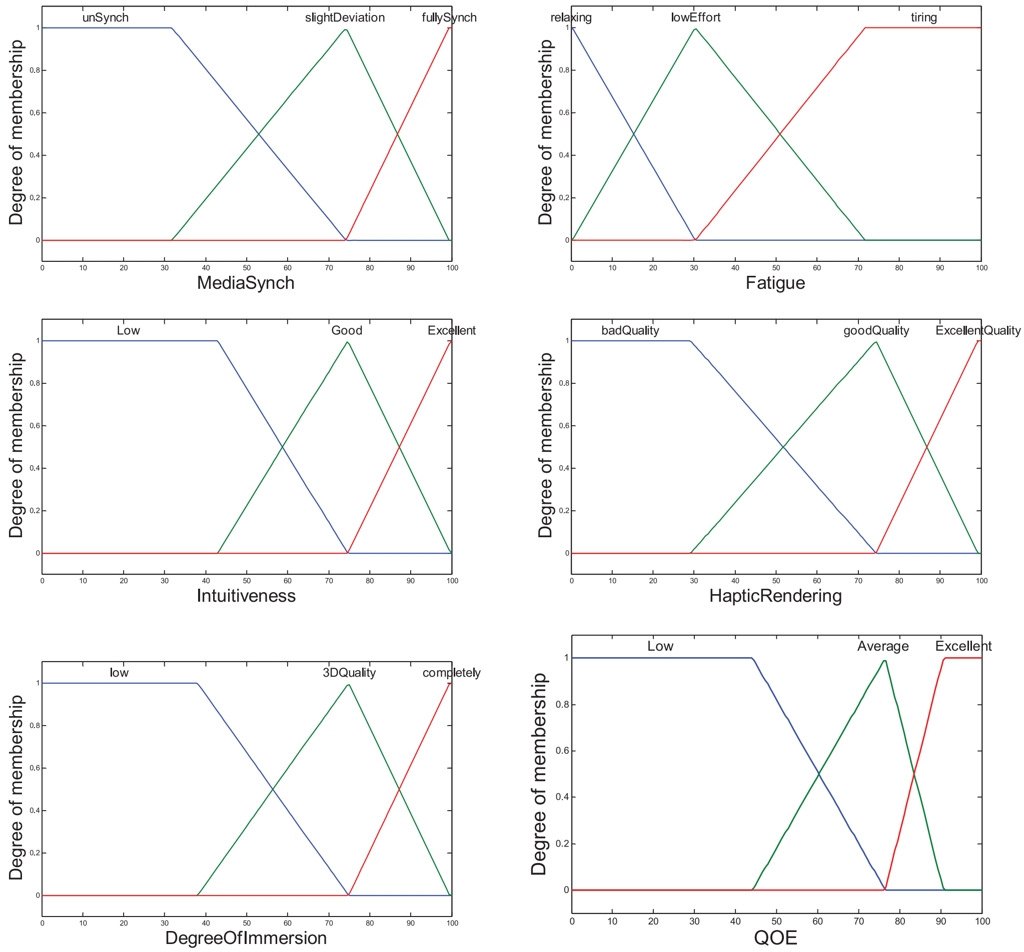

Fig. 10. Inputs/output fuzzy sets and membership functions.

Table II. Rule Conversion Example

	Media Synch	Fatigue	Rendering	Deg. of Immersion	User Intuitiveness	Overall Rating
U3	75	25	75	100	75	90
Rule	M	M	M	H	M	H

The rule generated from the 3rd user is translated as: if MediaSynch is medium (slight deviation) AND Fatigue is medium (low effort) AND Haptic Rendering is medium (good quality) AND Degree of Immersion is high (completely) AND User Intuitiveness is medium (good) THEN QoE is high (excellent).

Often, to generate the rule we have to resolve the value to its most probable linguistic MF. For example, if we took the input of fatigue in the above table and looked it up in the membership plot of fatigue (Figure 10), we will find that 25 on the x-axis has 0.18 truth value in the low MF and 0.82 truth value in the medium MF. The higher the truth value, the higher is the probability that the value

Table III. Testing the FIS

	Media Synch	Fatigue	Rendering	Deg. of Immersion	User Intuitiveness	Overall User Rating	FIS Output
U6	100	0	50	75	75	90	90.7
U12	75	0	100	75	100	95	89.5
U18	75	25	75	50	75	85	85.9
U24	50	25	50	50	50	65	72.9
U30	100	0	100	100	100	95	92

belongs to a membership function. Hence we considered 25 to belong to the medium MF. The rest of the rules were decided similarly (rules not listed for space limitation).

The first twenty five rules correspond to the twenty five users' data we used for clustering. Rules twenty six and twenty seven were additionally added. They essentially state that if all input parameters are high then QoE is high. Similarly if all inputs are low then QoE is low (except for the fatigue parameter which is reversed; meaning that low fatigue increases QoE while high fatigue decreases it).

4.6 Generating Output

Since our system is a Mamdani FIS, a defuzzification step is essential to get that crisp QoE output from the fuzzy aggregate that is generated after applying the rules to the input. The most common defuzzification method is the centroid calculation. This method calculates the center value of the aggregate fuzzy curve generated before the defuzzification step. The centroid method returns this value as the crisp output of the FIS.

5. TESTING AND ANALYSIS

5.1 Testing the Fuzzy Inference System

To test the fuzzy inference system, we ran the data from the rest of the users that were not used in the clustering procedure. The data and the outputs are presented in Table III.

The last column of the table displays the FIS output value when the five input values of the FIS are taken from the user. For example, when the data vector of user six [100 0 50 75 75] is given as input to the FIS, the output of the FIS is 90.7. The actual rating of QoE by user six is 90%, provided in the 'Overall User Rating' column.

5.2 Error Calculation

5.2.1 *Relative and Percent Error.* Observing the last two columns of Table III, we can compare the actual value of the user rating (QoE_u) and the FIS output (QoE_f). The average relative error can be computed by summing all the relative error and dividing by the number of users utilized for testing (n), according to the following equation:

$$re_{ave} = \frac{\sum_{i=1}^{n} |QoE_u - QoE_f| / QoE_u}{n} \tag{2}$$

$$re_{ave} = (0.0078 + .0579 + 0.0106 + 0.1215 + 0.0316)/5 = 0.04588.$$

Hence, the average relative error in the FIS output is 0.04588 and the percent error is approximately 4.6%.

5.2.2 *Root Mean-Square Error*. Another way to evaluate our FIS output is to consider the root mean-square error (RMSE). The RMSE could be calculated according to the following equation:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(QoE_u - QoE_f)^2}{n}}. \quad (3)$$

RMSE measures the difference between the predicted value of an estimator and the actual outcome. The idea is to calculate how far the FIS output is from the general category of the results. In our case RMSE = 4.538. This is an absolute and not a relative value. This indicates that on average the estimator (FIS) value deviates from the original value (user's value) by approximately 4.5 points on a scale of 100.

5.3 Correlation and Statistical Testing

In this section we provide a statistical analysis of the acquired results. First we calculated the correlation between the users' responses and the FIS output. The correlation value is 0.98 $p < 0.01$. The two sets of data correlates with each other significantly, indicating that the two groups are close in value to each other.

Yet, in this case we want to test whether the two sets of data are actually the same and that there is no significant difference between the FIS output and the users' evaluation. That is, we want to test that both the FIS output and the user evaluation represent the same group of users.

Unlike common statistical tests such as the Student's T-test represented by H1, where the researcher tries to prove the difference in the average of two groups (the alternate hypothesis HA of H1), our goal is to prove that there is no difference between the two sets of results (null hypothesis H0 of H1). In this case we can be significantly confident that the FIS output is a valid representation of the users' QoE.

$$\text{H0}: \mu 1 = \mu 2;\ \text{HA}: \mu 1 \neq \mu 2. \quad \text{(H1)}$$

To prove the equivalence of the two groups we followed the methodology called test of equivalence found in the literature [Streiner 2003; Cribbie et al. 2004]. The method consists of finding an acceptable difference (D) between both means of the results in which we still consider them equivalent. Hence the new null and alternate hypothesis become

$$\text{H01}: \mu 1 - \mu 2 > D;\ \text{H02}: \mu 1 - \mu 2 > -D;\ \text{HA}: \mu 1 - \mu 2 \leq D. \quad \text{(H2)}$$

Writing the null and alternate hypothesis in this way indicates that if the difference between the two means is greater than D then the two groups are different (null hypothesis) otherwise if the difference is less than or equal to D then the two groups are equivalent (alternate hypothesis HA of H2) which is what we want to prove. Note that the null hypothesis is divided into two parts to ensure any sign differences that may arise. In other words, both null hypotheses (H01 and H02 of H2) state that the difference between the two means resides outside the range of $[-D, D]$.

The t-value equations for rejecting each of the test of equivalence's null hypotheses and thus accepting the alternate hypothesis is

$$t_1 = \frac{(M_1 - M_2) - D}{S_{M_1 - M_2}} \quad (4)$$

$$t_2 = \frac{(M_1 - M_2) - (-D)}{S_{M_1 - M_2}}, \quad (5)$$

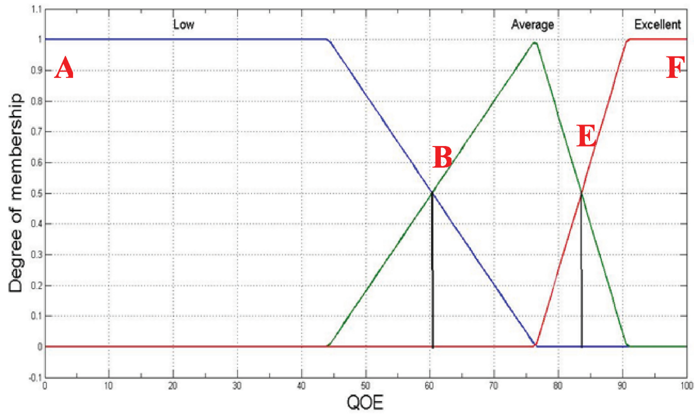

Fig. 11. Output MFs for Mamdani FIS with points of interest.

where M is the mean of the series, D is the acceptable difference, and $S_{M_1-M_2}$ is the standard error of the difference defined by:

$$S_{M_1-M_2} = \sqrt{\left[\frac{(n_1-1)\,s_1^2 + (n_2-1)\,s_2^2}{n_1+n_2-2}\right] x \left[\frac{1}{n_1} + \frac{1}{n_2}\right]}, \tag{6}$$

where s represents the standard deviation, n is the number of elements, and the degrees of freedom (DF) $= n_1 + n_2 - 2$.

In our case, $M1 = 86.0$, $S1 = 11.13$ for the column 'User Overall Rating', and $M2 = 86.2$, S2 $= 6.95$ for the column 'FIS output' (Table III). Therefore our standard error is $S_{M1-M2} = 5.87$.

Next we want to calculate the t-values. The challenge here is to find an acceptable D value. Observing the output function in Figure 11, there are four points of interest to consider: the two pole points of the graph (A and F) at $x = 0$ and $x = 100$ and the intersections of the MFs (B and E) that happen at $x = 60.5$ and $x = 83.5$.

If the FIS output is between the two intersection values (B and E) then the output would be resolved to the Average MF, since it will have the highest truth value (see deriving the rules of Section 5.3.4). If the output is above point E and below point F then it would be resolved to the Excellent MF. If the output is below point B and above point A then it would be resolved to the Low MF. Hence, to find a suitable D, we require that the output does not change its MF status (still to be resolved to the previous MF even if we added or subtracted difference D from the output). To find that D value, we subtracted the FIS output from the points of interest (intersection point or the pole point) of the MFs that would apply to it

$$|Output - applicable\ point\ of\ interest| = difference$$
$$|90.7 - 100| = 9.3$$
$$|90.7 - 83.5| = 7.2$$
$$|100 - 89.5| = 10.5$$
$$|89.5 - 83.5| = 6.0$$
$$|85.9 - 100| = 14.1$$

$$|85.9 - 83.5| = 2.4$$
$$|72.9 - 83.5| = 10.6$$
$$|72.9 - 60.5| = 12.4$$
$$|100 - 92.0| = 8.0$$
$$|92.0 - 83.5| = 8.5$$

The D value in this case would be the average of the differences. $D = 89/10 = 8.9$. Applying equations 4 and 5, we get

$$t_1 = \frac{(86 - 86.2) - 8.9}{5.87} = -1.55, \text{ and } t_2 = \frac{(86 - 86.2) - (-8.9)}{5.87} = 1.48$$

The critical value (tv) when $DF = 8$ is 1.397 for $p < 0.1$. According to Cribbie et al. [2004], if $t1 < -tv$ we can reject the null hypothesis H01 of H2. Moreover, if $t2 > tv$ we can reject the null hypothesis H02 of H2. We have $t1$ $(-1.55) < -1.397$ and $t2$ $(1.48) > 1.397$. Therefore we can reject the null hypotheses and accept the alternate hypothesis with $p < 0.1$. This suggests that the two groups are equivalent significantly with $p < 0.1$. This result is in agreement with the Student's t-test ($t = -0.0305$, $p = 0.98$) that indicates that we cannot reject the null hypothesis H0 of H1.

5.4 Discussion

For testing the FIS, we applied the following evaluation assessments: relative error, root mean error square, correlation, and statistical testing. Although the values are not identical between the user and the FIS output, the percentage error indicates that the FIS values are within a reasonable error value compared to the users' values (4.6%). The RMSE calculation confirms that the deviation between the FIS output and the users value is less than five, which on a scale of 100 is moderately low.

Correlation value between the two results was high as expected given that the results generated low relative and percent error values. The high correlation value is significant, $p < 0.01$, which indicates that the two groups of values follow closely the same pattern. The correlation values are important in detecting the direction of the results, but in this case we wanted to prove that the FIS values and the users' ratings represented the same group and that there is no significant difference between the two.

We resorted to two types of statistical analysis which complemented each other. While the Student's t-test produced a very high p value indicating that we should reject the alternate hypothesis HA of H1, that did not automatically indicate that we can accept the null hypothesis H0 of H1 (although it indicates that accepting the null hypothesis would still be an option), as statisticians suggest. This is the reason we performed the second type of statistical analysis, the test of equivalence. With a calculated difference D, we were able to accept the alternate hypothesis HA of H2, $p < 0.1$. The alternate hypothesis states that both results represent the same group within a small range of $[-D, D]$.

6. CONCLUSION

This article has introduced a taxonomy for classifying QoE parameters. The taxonomy was based on the definition of the QoE and was divided into QoS and UX groups. Each group was subdivided into certain measures that falls under that category.

Moreover, this article has presented a fuzzy logic inference system capable of evaluating the user's QoE based on certain input parameters. The FIS results mimic the users' ratings with low relative and root mean-square error. FIS was chosen due to the vagueness of the inputs as well as the nonlinearity of the input/output relation. The set of rules derived map the input parameters to a crisp QoE value.

The QoE value obtained represents an alternate to the user's rating since the results obtained are significant with high correlational values. Hence, by validating the FIS results, we can use these systems as bases for alternatives to users' subjective tests in the future.

APPENDIX: USER STUDY QUESTIONNAIRE

1. To what extent did the game provide an immersive environment in which you were completely engaged?

   Not immersive — Completely immersive

   1 2 3 4 5

2. To what extent do you think the haptic playback was realistic?

   Not realistic — Completely realistic

   1 2 3 4 5

3. To what extent do you think the haptic feedback was useful?

   Not useful — Completely useful

   1 2 3 4 5

4. To what extent, if any, did using the haptic device increase interaction intuitivism?

   Not at all — Completely

   1 2 3 4 5

5. To what extent, if any, did using the haptic device cause fatigue?

   Not at all — Completely

   1 2 3 4 5

6. To what extent did you have a sense of being with the other person?

   Not at all — Completely

   1 2 3 4 5

7. How would you rank your experience using the haptic device?

   Very Poor — Very Good

   1 2 3 4 5

8. How would you rank the synchronization between graphics and haptic feedback?

   Very Poor — Excellent

   1 2 3 4 5

How familiar are you with haptic devices? (Please circle one)

1 What is a haptic device?
2 Familiar with them but never used them
3 Have used them before but only once or twice
4 Use them regularly
5 Have programming and/or engineering knowledge of them

Give a grade, over 100, for the overall quality of the application? /100

Gender: Male Female

Thank you for participating in this study!

REFERENCES

H. Al Osman, M. Eid, and A. El Saddik. 2008. Evaluating ALPHAN: A network protocol for haptic interaction. In *Proceedings of the 16th Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems*.

A. Alamri, M. Eid, and A. El Saddik. 2008. A quality of performance model for evaluating post-stroke patients. In *Proceedings of the IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA'08)*. 14–18.

C. Basdogan, C. H. Ho, M. A. Srinivasan, and M. Slater. 2000. An experimental study on the role of touch in shared virtual environments. *ACM Trans. Comput.-Hum. Interact.* 7, 443–460.

K. Behr, A. Nosper, C. Klimmt, and T. Hartmann. 2005. Some practical considerations of ethical issues in VR research. *Presence: Teleoperators Virtual Environ.* 14, 668–676.

A. Bhattacharya, W. Wu, and Z. Yang. 2011. Quality of experience evaluation of voice communication systems using affect-based approach. In *Proceedings of the 19th ACM International Conference on Multimedia.* 929–932.

K. Black. 2004. *Business Statistics: For Contemporary Decision Making.* Leyh Publishing.

R. A. Cribbie, J. A. Gruman, and C. A. Arpin-Cribbie. 2004. Recommendations for applying tests of equivalence. *J. Clinical Psych.* 60, 1–10.

F. Danieau, J. Fleureau, A. Cabec, P. Kerbiriou, P. Guillotel, N. Mollet, M. Christie, and A. L'Ecuyer. 2012. Framework for enhancing video viewing experience with haptic effects of motion. In *Proceedings of the Haptics Symposium*. 541–546.

L. De Marez and K. De Moor. 2007. The challenge of user- and QoE-centric research and product development in today's ICT-environment. *Observatorio (OBS*)* 1, 1–22.

C. De Vleeschouwer, T. Delmot, X. Marichal, and B. Macq. 1997. A fuzzy logic system for content-based bit-rate allocation. *Signal Proces. Image Commun.* 10, 115–141.

T. Ebrahimi. 2009. Quality of multimedia experience: past, present and future. In *Proceedings of the 17th ACM International Conference on Multimedia*. 3–4.

A. El Saddik. 2007. The potential of haptic technologies. *IEEE Instrum. Meas. Mag.* 10, 1, 10–17.

A. Gaggioli, M. Bassi, and A. Fave. 2003. Quality of experience in virtual environments. *Emerging Commun.* 5, 121–136.

A. Guerraz, C. Loscos, and R. Widenfeld. 2003. How to use physical parameters coming from the haptic device itself to enhance the evaluation of haptic benefits in user interface? In *Proceedings of the Eurohaptics Conference*. 137–145.

F. Gutierrez, J. Pierce, V. M. Vergara, R. Coulter, L. Saland, T. P. Caudell, T. E. Goldsmith, and D. C. Alverson. 2007. The effect of degree of immersion upon learning performance in virtual reality simulations for medical education. *Stud. Health Tech. Informatics* 125, 155–160.

I. Hajshirmohammadi and S. Payandeh. 2007. Fuzzy set theory for performance evaluation in a surgical simulator. *Presence: Teleoperators Virtual Environ.* 16, 603–622.

A. Hamam, M. Eid, A. E. Saddik, and N. D. Georganas. 2008a. A quality of experience model for haptic user interfaces. In *Proceedings of the Ambi-Sys Workshop on Haptic User Interfaces in Ambient Media Systems (HAS'08)*. 1–6.

A. Hamam, M. Eid, A. E. Saddik, and N. D. Georganas. 2008b. A fuzzy logic system for evaluating quality of experience of haptic-based applications. In *Proceedings of the Eurohaptics Conference*. Springer, 129–138.

A. Hamam, M. Eid, and A. E. Saddik. 2013. A. Effect of kinesthetic and tactile haptic feedback on the quality of experience of edutainment applications. *Multimedia Tools Appl.* 67, 2, 455–472.

A. Hamam and El Saddik, 2013. Towards a mathematical model for quality of experience evaluation of haptic applications. *IEEE Trans. Instrum. Meas.* 62, 12. DOI: 10.1109/TIM.2013.2272859.

P. Huang, Y. Ishibashi, N. Fukushima, and S. Sugawara. 2012. QoE assessment of olfactory media in remote ikebana with haptic media. In *Proceedings of the IEEE International Workshop Technical Committee on Communications Quality and Reliability*. 1–6.

Y. Ida, Y. Ishibashi, N. Fukushima, and S. Sugawara. 2010. QoE assessment of interactivity and fairness in First Person Shooting with group synchronization control. In *Proceedings of the Annual Workshop on Network and Systems Support for Games*. 1–2.

W. Ijsselsteijn, Y. De Kort, K. Poels, A. Jurgelionis, and F. Bellotti. 2007. Characterising and measuring user experiences in digital games. In *Proceedings of the International Conference on Advances in Computer Entertainment Technology*.

K. Iwata, Y. Ishibashi, N. Fukushima, and S. Sugawara. 2010. QoE assessment in haptic media, sound, and video transmission: Effect of playout buffering control. *Comput. Entertain.* 8, 12:1–12:14.

R. Jain. 2004. Quality of experience. *IEEE Multimedia* 11, 96–95.

L. A. Jones and H. Ho. 2008. Warm or Cool, Large or Small? The Challenge of Thermal Displays. *IEEE Trans. Haptics* 1, 53–70.

M. Karam, F. A. Russo, and D. I. Fels. 2009. Designing the Model Human Cochlea: An Ambient Crossmodal Audio-Tactile Display. *IEEE Trans. Haptics* 2, 160–169.

Y. Kusunose, Y. Ishibashi, N. Fukushima, and S. Sugawara. 2010. QoE assessment in networked air hockey game with haptic media. In *Proceedings of the Annual Workshop on Network and Systems Support for Games*. 1–2.

K. S. Park and R. V. Kenyon. 1999. Effects of network characteristics on human performance in a collaborative virtual environment. In *Proceedings of the IEEE Conference on Virtual Reality*.

E. H. Mamdani and S. Assilian. 1999. An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. Human-Computer Studies* 51, 135–147.

MATLAB Documentation, Fuzzy logic toolbox for use with MATLAB. Math Works Inc., 2001.

V. G. Miller. 1993. Measurement of self-perception of intuitiveness. *Western J. Nursing Res.* 15, 595.

C. C. Preston and A. M. Colman. 2000. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica 104*.

L. Nacke and C. A. Lindley. 2008. Flow and immersion in first-person shooters: measuring the player's gameplay experience. In *Proceedings of the Conference on Future Play: Research, Play, Share*. ACM, 81–88.

L. E. Nacke and C. A. Lindley. 2010. Affective ludology, flow and immersion in a first-person shooter: Measurement of player experience. *Loading* 3, 5.

A. Ramsey. 1997. Investigation of physiological measures relative to self-report of virtual reality induced sickness and effects (VRISE). In *Proceedings of the International Workshop on Motion Sickness: Medical and Human Factors*. 26–28.

G. H. Roid. 2004. Quality of performance and change-sensitive assessment of cognitive ability. In *Proceedings of the International Test Users Conference*. Austrahan Conference for Educational Research.

L. A. Rowe and R. Jain. 2005. ACM SIGMM retreat report on future directions in multimedia research. *ACM Trans. Multimedia Comput. Commun. Appl.* 1, 3–13.

R. Seroussi, M. H. Krag, P. Wilder, and M. H. Pope. 1989. The design and use of a microcomputerized real-time muscle fatigue monitor based on the medical frequency shift in the electromyographic signal. *IEEE Trans. Biomed. Eng.* 36, 284–286.

M. A. Srinivasan and C. Basdogan. 1997. Haptics in virtual environments: Taxonomy, research status, and challenges. *Computers Graphics* 21, 393–404.

D. L. Streiner. 2003. Unicorns do exist: a tutorial on "proving" the null hypothesis. *Canadian J. Psych.* 48, 756–761.

D. Strickland, L. Hodges, M. North, and S. Weghorst. 1997. Overcoming phobias by virtual exposure. *Commun. ACM* 40, 34–39.

P. Sweetser and P. Wyeth. 2005. GameFlow: a model for evaluating player enjoyment in games. *Computers Entertain.* 3, 3–3.

A. Tatematsu, Y. Ishibashi, N. Fukushima, and S. Sugawara. 2010. QoE assessment in haptic media, sound and video transmission: Influences of network latency. In *Proceedings of the IEEE International Workshop Technical Committee on Communications Quality and Reliability*. 1–6.

L. X. Wang and J. M. Mendel. 1992. Generating fuzzy rules by learning from examples. *IEEE Trans Syst. Man Cyber.* 22, 1414–1427.

T. E. Whalen, S. Noel, and J. Stewart. 2003. Measuring the human side of virtual reality. In *Proceedings of the IEEE International Symposium on Virtual Environments, Human- Computer Interfaces and Measurement Systems (VECIMS'03)*. 8–12.

W. Wu, A. Arefin, R. Rivas, K. Nahrstedt, R. Sheppard, and Z. Yang. 2009. Quality of experience in distributed interactive multimedia environments: toward a theoretical framework. In *Proceedings of the 17th ACM International Conference on Multimedia*. 481–490.

W. Wu, A. Arefin, Z. Huang, P. Agarwal, S. Shi, R. Rivas, and K. Nahrstedt. 2010. "I'm the Jedi!" - A case study of user experience in 3D tele-immersive gaming. In *Proceedings of the IEEE International Symposium on Multimedia.* 220–227.

Received September 2012; revised May 2013; accepted September 2013